Received: 12 February 2020 | Accepted: 12 April 2020
DOI: 10.1002/mpr.1829

SPECIAL ISSUE

The Saudi National Mental Health Survey: Sample design and weight development

Zeina N. Mneimneh[1] | **Steven G. Heeringa**[1] | **Yu-Chieh Lin**[1] | **Yasmin A. Altwaijri**[2,3,4] | **Raphael Nishimura**[1]

[1]Institute for Social Research, University of Michigan, Ann Arbor, Michigan

[2]King Salman Center for Disability Research, Riyadh, Saudi Arabia

[3]Biostatistics, Epidemiology and Scientific Computing Department, King Faisal Specialist Hospital and Research Centre—Riyadh, Riyadh, Saudi Arabia

[4]SABIC Psychological Health Research & Applications Chair (SPHRAC), College of Medicine, King Saud University, Riyadh, Saudi Arabia

Correspondence
Zeina N. Mneimneh, Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48104.
Email: zeinam@umich.edu

Funding information
Abraaj Capital; King Abdulaziz City for Science and Technology (KACST); King Faisal Specialist Hospital and Research Center; King Saud University; Ministry of Economy and Planning, General Authority for Statistics; Ministry of Health (Saudi Arabia); Saudi Basic Industries Corporation (SABIC)

Abstract

Objectives: To describe the sample design and weighting procedures used in the Saudi National Mental Health Survey (SNMHS).

Methods: A multistage clustered area probability design was used to select the SNMHS sample with one male and one female KSA citizen ages 15–65 surveyed in each sample household.

Results: A design representative of the household population was developed and modified iteratively to adjust for unanticipated field complications. These modifications, along with variation in within-household probabilities of selection and geographic–demographic variation in response rates were accounted for through survey weights. Design-based estimation methods were used to adjust for the effects of these weights and of geographic clustering. Design effects were estimated and simulations were carried out on bias-variancetrade-offs in weight trimming to evaluate the implication of design features for precision of estimates.

Conclusions: The multiple purposes of the survey will require the use of different weights for different types of analyses, including household and person weights as well as weights for proxy reports about household members whose disabilities prevented them from participating in the survey. It will be important to use these different weights appropriately in the diverse analyses that will be undertaken with the SNMHS data.

KEYWORDS
Composite International Diagnostic Interview (CIDI), design effects, Saudi National Mental Health Survey (SNMHS), WHO World Mental Health (WMH) Survey Initiative

1 | INTRODUCTION

This article presents a description of the sample design and weighting procedures used in the Saudi National Mental Health Survey (SNMHS). The SNMHS is a nationally representative household survey of the prevalence and correlates of common mental disorders in the Kingdom of Saudi Arabia (KSA). Prior to the SNMHS, only limited data were available on the burden of mental disorders in KSA. The Global Burden of Disease (GBD) Study estimated that drug use disorders, depressive disorders, and anxiety disorders are the third, fourth, and sixth leading causes of disability in KSA (Institute for Health Metrics and Evaluation, 2019). However, these estimates were based on extrapolations using indirect information from epidemiological surveys carried out in other countries in the region (GBD 2015 Eastern Mediterranean Region Mental Health Collaborators, 2018). Given the important policy implications of these estimates, more direct data are needed for policy planners to assess the societal burden of mental disorders, unmet need for treatment, and barriers to treatment.

The SNMHS was launched to provide these data as part of the World Health Organization (WHO) World Mental Health (WMH) Surveys Initiative (Alonso, Chatterji, & He, 2013; Kessler & Üstün, 2008; Scott, de Jonge, Stein, & Kessler, 2018). Standardized WMH methods were used in SNMHS field implementation to provide valid data on the prevalence and distribution of mental disorders and unmet need for treatment of these disorders (Harkness et al., 2008; Heeringa et al., 2008; Pennell et al., 2008). However, as detailed in a prior article in this issue (Al-Subaie et al., In press), several field implementation procedures were changed in order to adjust the design to the special circumstances of KSA. The sample design was also changed in order to use the survey to study several issues of special policy importance in KSA. These changes, in turn, led to modifications of the standard WMH weighting procedures. We present a broad overview of the design and discuss these special features in this article.

The SNMHS is a project of the King Salman Center for Disability Research in collaboration with the King Faisal Specialist Hospital and Research Centre, the Saudi Ministry of Health, King Saud University, the Ministry of Economy and Planning, and the General Authority for Statistics. Survey design and implementation and data analysis support are provided by the WMH Data Collection and Data Analysis Coordination Centers at the University of Michigan Survey Research Center (SRC) and Harvard Medical School (HMS), respectively. The survey was conducted between 2011 and 2016. This timeline includes interruptions in the fieldwork because of delay in receiving funding for the survey and change in data collection agency in 2013.

2 | SAMPLE DESIGN AND SELECTION

The SNMHS was initially designed to be nationally representative of Saudi citizens between the ages of 15 and 65 years living in urban and rural areas. However, due to security concerns resulting from the ongoing armed conflicts with neighboring countries and limited access to remote areas, two out of the 13 administrative areas, Jazan and Najran, were excluded from the survey population. A stratified multistage cluster area probability sample of the noninstitutionalized Saudi citizens (15–65 years old) was selected from the remaining 11 administrative areas: Riyadh, Makkah, Al-Madinah, Al Qaseem, Eastern Province, Aseer, Tabouk, Hail, Northern Region, Al-Baha, and Al-Jouf. The survey population was stratified by these 11 administrative areas. In each of these 11 strata, the primary sampling units (PSUs) were the census count administrative areas, defined as per maps provided and updated by the Ministry of Economy and Planning (General Authority for Statistics, 2010). After establishing a minimum number of PSUs to be selected in each of the smaller strata, the rest of the PSUs were allocated approximately proportionate to the number of Saudi households in the population of each stratum according to the 2010 Census.

A probability sample of PSUs in each stratum was then selected in multiple iterations in collaboration with the General Authority for Statistics (GaStat) of KSA using their national frame of PSUs. The PSUs were first sorted by location and size (main cities with population over 100,000, urban cities with population between 5,000 and 100,000, and villages with population below 5,000). In this first iteration, GaStat selected a sample of 473 PSUs with a systematic probability proportional to size (PPS) sampling, in which the measure of size was the total number of households (Saudi and non-Saudi) according to the 2010 Census. Due to operational and cost constrains, the total number of PSUs was reduced. In this iteration, a subsample of 404 PSUs was selected with systematic probabilities proportional to size, using as the measure of size the number of Saudi households. While the initial 473 PSUs was selected based on the total number of households in each PSU, the subsample of 404 was selected based on the total number of Saudi households which was not available at the time of the initial selection.

Early monitoring of the field cost indicated that the study could not afford to conduct the survey in the larger sample of 404 PSU locations. To control costs, the PSU sample size was reduced further. To compensate for the reduction in numbers of sample PSUs, the total size of the sample was maintained by increasing the targeted number of sample households within each PSU cluster (see below). This third step in the final determination of the sample of PSUs employed two rules. First, PSUs that contained completed interviews were retained in the final sample with certainty. Second, within each primary stage sample stratum, sample PSUs that had not yet been released to the field were subsampled with equal probability.

The second stage of sampling involved the selection of households. The household frame for each selected PSU consisted of the list of addresses compiled during the 2010 Census conducted by GaStat. Households, therefore, were sampled by selecting addresses, assuming a one-to-one correspondence. For each selected PSU, 38 addresses were selected randomly. The sample of 38 addresses for each PSU was divided into random subsamples or "replicates" to ensure control over final sample size as the field period progressed. Due to the operational and costs constraints mentioned earlier, not all random replicates of the sample addresses were released for screening and interviewing in some of the PSUs. Thus, the probability subsample of selected addresses within each PSU ranged from 16 to 38. Only residential addresses were considered eligible.

The third stage of sampling consisted of selecting up to two random household members (a male and a female) within each selected address. Each address was assigned to an interviewer to complete a household roster with an informant who was a household member, determine the eligibility of each household member, and ask the informant about any physical or mental disability for each listed member. Eight different types of disabilities were assessed: serious hearing, vision, or speech impairment; missing limb or paralysis; other physical impairment, disfigurement, or handicap; other serious chronic physical illness or pain; temporary physical illness or disability; concentration, memory, or decision making problems; learning disability, down syndrome, or mental retardation; and serious mental problem. Eligibility criteria included being a Saudi citizen between the ages of 15 and 65 years and being able to speak Arabic. Once eligible members were identified, one random female and one random male were selected within each gender group. Eligible members with at least one reported disability were over-sampled by a factor of two (i.e., two times the chance of other household members of being selected). Random selection of designated respondents from the household rosters of eligible men and women was automated and conducted by an algorithm programmed to be part of the computer-assisted personal interview (CAPI) instrument so as to avoid the possibility that an interviewer would try to select an easy-to-reach household member or substitute a cooperative household member for an originally selected member that was not cooperative.

Table 1 summarizes the final number of PSUs retained in the sample, the number of households selected, the number of respondents, and the final number of completed interviews per stratum (Table 1). The overall sample included selection of 4,302 households. The household screening rate was 84% and the conditional interview response rate was 73%, for an estimated individual-level response rate of 61% (American Association for Public Opinion Research, 2016). As detailed in the previous paper (Al-Subaie et al., In press), the noncore sections were administered to a subsample of respondents as a design feature to reduce the overall interview length.

3 | WEIGHTING

Several weights were developed for the SNMHS to enable estimation and inference for several types of populations.

3.1 | Household level weights

3.1.1 | First-stage weights

Base weights were calculated considering the probability of selection in each of the sampling stages within each stratum, defined by the administrative areas, as previously described. The PSUs of the multistage sample are enumeration area units as defined by the KSA population Census. The selection probability of the PSUs accounts for three distinct steps in this first sampling stage:

1. GaStat of KSA assisted in the first step of selecting the PSUs. Using the KSA frame of enumeration areas, GaStat performed an initial PPS selection of 473 PSUs across the 11 administrative areas, using total number of households as the measure of size:

$$f_{1h\alpha} = a_{1h} \frac{\mathrm{MoS}_{h\alpha}}{\sum_{\alpha=1}^{A_h} \mathrm{MoS}_{h\alpha}},$$

where A_h is the total number of PSUs in stratum h, a_{1h} is the initial number of PSUs selected in stratum h, $\mathrm{MoS}_{h\alpha}$ is the total number of households in PSU α within stratum h. GaStat staff reviewed the census data for each of the 473 selected PSUs and provided the study team with a data set that included PSU identifiers, total household measures of size and a count of the number of households that included Saudi citizens.

2. Using the data on the selected 473 PSUs, the study team conducted a second sampling step, converting the original total household measures of size to a more efficient sample based on eligible Saudi household measures of size. In this second step, a reduced

TABLE 1 The Saudi National Mental Health Survey sample design

Stratum	% of Saudi population (census)	# of PSUs selected	# of households selected	# of respondents selected	# of completed interviews
Riyadh	24.99%	41	1,268	1,432	880
Makkah	24.74%	53	824	1,120	925
Al-Madinah	7.01%	12	230	312	241
Al Qaseem	5.33%	10	294	400	271
Eastern Province	16.98%	18	371	682	518
Aseer	9.10%	15	375	410	310
Tabouk	3.51%	8	248	230	163
Hail	2.88%	5	165	182	133
Northern Frontier	1.47%	5	80	105	94
Al-Baha	2.07%	5	135	175	146
Al-Jouf	1.93%	12	312	414	323
Total	100.00%	184	4,302	5,462	4,004

Abbreviation: PSUs, primary sampling units.

sample of 404 PSUs of the 473 initial PSUs was subselected with PPS, but using the number of Saudi households as the measure of size:

$$f_{2h\alpha} = a_{2h}\frac{M_{h\alpha}}{\sum_{\alpha=1}^{a_{1h}} M_{h\alpha}},$$

where a_{2h} is the number of PSUs subsampled in stratum h, and $M_{h\alpha}$ is the total number of Saudi households in PSU α within stratum h.

3 As noted above, a second iteration of PSU subsampling was implemented due to cost constraints after the initial release of PSUs in several strata that represented the large metropolitan areas. PSUs that had already been worked were retained in the final sample with certainty. Within each primary stage sample stratum, sample PSUs that had not yet been released to the field were subsampled with equal probability. Based on this subsampling plan, a final probability sample of 184 PSUs was retained for the study data collection. The conditional probability that an original sample PSU was retained in this third cost-saving subsampling step is:

$$f_{3h\alpha} = \begin{cases} 1, \text{if PSU } \alpha \text{ had already been worked} \\ \frac{a_{3h}}{a_{2h}}, \text{otherwise} \end{cases},$$

where a_{3h} is the number of PSUs subsampled in stratum h during data collection.

Across all three steps employed in selecting the final sample of PSUs, the first-stage selection probability for PSU α within stratum h was then calculated as

$$f_{h\alpha} = f_{1h\alpha} \times f_{2h\alpha} \times f_{3h\alpha}$$

3.1.2 | Second-stage weights

In the second sampling stage, households within the selected PSUs were sampled with equal probability. Therefore, the selection probability for household β, conditional to the selection of PSU α within stratum h was

$$f_{\beta|h\alpha} = \frac{b_{h\alpha}}{M_{h\alpha}},$$

where $b_{h\alpha}$ is the number of households selected in PSU α within stratum h.

The household base weight for a household β in PSU α within stratum h was then given by the inverse of the product of the sampling rates across these first two stages:

$$f_{h\alpha\beta} = f_{h\alpha} \times f_{\beta|h\alpha}$$

In order to address potential nonresponse bias, a household nonresponse adjustment was computed at two levels: (a) the screening interview and (b) the main interview. At both levels, the nonresponse adjustment factors were computed as the inverse of the respective response rates within the PSU.

The final household weight was then computed as the product of its base weight and nonresponse adjustment factors.

3.2 | Respondent level weights

As noted above, up to two eligible individuals were randomly selected within each selected household, one female and one male. Eligible individuals with any reported disability were over-sampled by a factor of two. To account for the unequal probability of selecting a respondent, a base weight was calculated as follows:

The probability of selection of an individual γ, conditional to the selection of household β in PSU α within stratum h is calculated as:

$$f_{\gamma|h\alpha\beta} = \begin{cases} \frac{1}{c_{h\alpha\beta}}, \text{if individual is male} \\ \frac{1}{d_{h\alpha\beta}}, \text{if individual is female} \end{cases}$$

where $c_{h\alpha\beta}$ and $d_{h\alpha\beta}$ are, respectively, the total number of eligible males and the total number of eligible females in a household β of PSU α within stratum h that does not contain any eligible individuals with a reported disability. In households with at least one individual with a reported disability, the calculation of $f_{\gamma\mid h\alpha\beta}$ is modified to reflect the fact that those individuals have twice the chance of being selected compared to the other eligible household members.

The overall selection probability for an eligible individual γ in household β within PSU α and stratum h is then the product of the household and the individual sampling probabilities:

$$f_{h\alpha\beta\gamma} = f_{h\alpha} \times f_{\beta|h\alpha} \times f_{\gamma|h\alpha\beta}$$

The individual base weight for Part I respondents (who were given the core sections) is computed as the inverse of this selection probability.

The individual base weight for Part II respondents (who were given the noncore sections) was adjusted by a factor that accounts for the probability of being selected to be part of the Part II sample. The Part II sample came from two pools of respondents. Those who endorsed a mental disorder during the core sections and who were selected with certainty (i.e., assigned a factor of 1) and those who did not endorse a mental disorder during the core sections and who were selected with a probability of 0.25 to Part II of the instrument (i.e., assigned a factor of 4).

Similar to the household weight, in order to mitigate for potential nonresponse bias, a nonresponse adjustment was performed at the PSU level. Within each PSU, the nonresponse adjustment factor was computed as the inverse of the PSU response rate by gender.

The joint product of the individual base weight (for Part I and for Part II separately) and the nonresponse adjustment factors at the household and individual levels were consolidated for each respondent.

The last component of weighting was made to account for differences between the weighted sample distribution (after the product of design and nonresponse weight was applied) and the general Saudi population distribution according to the 2010 Census data on various auxiliary variables. Such adjustment is generally known as calibration. If the auxiliary variables used in the adjustment are associated with the survey outcomes, calibration can decrease biases due to nonsampling errors and improve the precision of the survey estimates. This calibration procedure was implemented through a technique known as post-stratification, in which the sample is stratified according to auxiliary variables, and the weights are adjusted such that the distribution of the sample in those strata match the population distribution. This post-stratification adjustment was based on the general Saudi population distribution by gender, age, and region, considering only Saudi citizens aged 15–65 years old in KSA, and ensuring that the joint distribution of gender, age, and region in the weighted sample matches the known Saudi population joint distribution.

The final respondent weights for Part I and Part II were then normalized to ensure that the sum of the weights was equal to the total sample size of the Part I or Part II data sets. Table 2 shows the Part I and Part II sample distributions by gender and age, weighted and unweighted, as well as the 2010 Census Saudi population distributions. Comparison of these distributions provides information on the effects of weighting (Table 2). As shown, the unweighted Part I and Part II samples slightly overrepresent females. These distributions were corrected with the consolidated Part I weight for the Part I sample, and with Part II weight for the Part II sample.

3.3 | Special weights for proxy reports about household members with disabilities

Two sections of the core questionnaire, Physical Disability and Dementia, collected information on functional limitations due to the different types of disability reported by respondents concerning their own disabilities. Informant versions of these two sections were also created to allow respondents to serve as proxy respondents for household members who were excluded from being interviewed based on the reports from the initial household informant during the listing phase. In order to obtain population estimates using the Self and Proxy information from these sections, an additional weight was created and assigned to every household member, eligible for those sections or not. This weight was computed as the final household weight post-stratified to the general Saudi population distribution by gender, age, and region, considering only Saudi citizens in KSA. This adjustment was made considering every household member because the population distribution only for the eligible cases for these sections is unknown, whereas the population distribution for the overall Saudi population is known. While this weight was assigned to each household member, the proxy analysis of these two sections is being restricted to the individuals for whom the proxy information was collected in those sections.

4 | DESIGN-BASED ESTIMATION

Complex sample design features, such as stratification, clustering, and unequal probabilities of selection, introduce variability in the precision of survey estimates that is not taken into account by conventional statistical methods that assume simple random sampling (SRS). In order to

TABLE 2 Demographic distribution of sample compared to the census Saudi population on post-stratification variables

	Part I sample unweighted (%)	Part II sample unweighted (%)	Part I sample weighted (%)	Part II sample weighted (%)	Census (%)
Gender					
Male	47.4	42.1	50.5	50.5	50.5
Female	52.6	57.9	49.5	49.5	49.5
Age					
15–19	15.1	16.0	17.0	17.0	17.0
20–29	26.2	28.3	30.4	30.4	30.4
30–39	27.1	29.4	23.1	23.1	23.1
40 and above	31.6	26.4	29.5	29.5	29.5
Region[a]					
Central	28.7	30.1	30.5	28.3	30.5
Southern	11.4	11.9	10.9	12.2	10.9
Northern	17.8	15.0	9.8	7.9	9.8
Western	29.1	29.5	31.6	34.2	31.6
Eastern	12.9	13.5	17.3	17.4	17.3

[a]Regions categorized as per following strata: Central = Riyadh, Al Qaseem; Southern = Al-Baha, Aseer; Northern = Tabouk, Hail, Northern Frontier, Al-Jouf; Western = Makkah, Al-Madinah; Eastern = Eastern Province.

account for such features, design-based methods for estimating standard errors are used in the analysis of SNMHS data. The Taylor series linearization method (Wolter, 1985) is the approach we use to estimate standard errors of simple descriptive statistics in a way that accounts for the sample design characteristics. Simulations based on pseudo-sampling within the sample are used to obtain the standard errors of more complex statistics. These methods are described in detail elsewhere (Heeringa, West, & Berglund, 2017). The important feature of both methods for current purposes, though, is the need to use stratification and clustering information for each respondent to calculate these estimates. This was done by creating pseudo strata and pseudo PSUs.

As noted earlier, the study uses a sample design with PSUs selected from 11 administrative regions. Thirty-four standard error strata (SE-strata) were created proportionally based on the number of PSUs in each of these regions. The PSUs were assigned to SE-strata based on sample frame selection order. Within each of the SE-strata, all PSUs were randomized into one of the two standard error clusters (SE-cluster = 1 or SE-cluster = 2). After assigning the randomized values to a given SE-cluster, a cross tabulation was made between SE-strata and SE-cluster for both the Part I and Part II samples to check that the random PSU group defined by the SE-cluster did not contain a small sample (less than 10). Any small sample found was then manually rebalanced by modifying the SE-cluster (from 1 to 2 or 2 to 1).

The effect of clustering and weighting is measured by the design effect, which is defined as the ratio of the design-based sampling variance of a survey estimate and its corresponding SRS sampling variance (Kish, 1965). Both clustering and weighting tend to decrease the precision of survey estimates compared to SRS, leading to the design effect being larger than 1.0. The design effect tends to be somewhat smaller for multivariate analyses (e.g., estimates of regression coefficients) than for univariate analyses (e.g., prevalence estimates). For this reason, we restrict here the evaluation of the design effect to prevalence estimates of mental disorders that are assessed at the respondent level. Table 3 presents the design effect for various prevalence estimates of mental disorders using the Part I weight if the disorder was assessed in a core section and the Part II weight if the disorder was assessed in a noncore section.

With the exception of 12-month alcohol abuse, the design effects of all prevalence estimates are larger than 1, indicating some losses in precision compared to SRS. This means that conventional statistical methods would tend to under-estimate these standard estimates and in the case of association, consider some estimates statistically significant that would not be judged significant using the correct test. The largest design effect is for 12-month subthreshold bipolar II, where the design effect is 2.1, indicating a twofold loss of precision compared to an SRS design. Most of the other prevalence estimates have a design effect smaller than 1.5.

We investigated the value of using a weight trimming strategy that trades off bias for efficiency by trimming extreme weights. It sometimes occurs that this approach, although increasing bias, will lead to a reduction in total estimation error to the extent that the increase in bias is more than counter-balanced by a decrease in inefficiency due to the variance introduced by weighting. This has sometimes been the case in other WMH surveys (Heeringa et al., 2008; Kessler et al., 2013; Kessler, Heeringa, Pennell, Sampson, & Zaslavsky, 2018). However, the results of such an analysis applied to the SNMHS data (not shown here) indicate that trimming extreme weights does not reduce total survey error in estimating prevalence estimates. That is, the reduction in inefficiency by reducing weights was not found to be greater than the increase in bias. Based on this result, the final SNMHS weights were not trimmed.

TABLE 3 Design effects lifetime and 12-month disorder prevalence estimates

	Lifetime	12-month
Anxiety disorders		
Agoraphobia[a]	1.6	1.7
Social phobia[a]	1.7	1.6
Generalized anxiety disorder[a]	1.6	1.4
Post-traumatic stress disorder[b]	1.1	1.1
Obsessive–compulsive disorder[b]	1.5	1.1
Separation anxiety disorder[b]	1.9	1.3
Mood disorders		
Major depressive episode[a]	1.8	2.0
Bipolar I[a]	1.5	1.6
Bipolar II, subthreshold[a]	1.7	2.1
Disruptive behavior disorders		
Conduct disorder[b]	1.0	1.0
Attention-deficit/hyperactivity disorder[b]	1.8	1.3
Intermittent explosive disorder[b]	1.3	1.5
Substance disorders		
Alcohol abuse[b]	1.4	0.6
Alcohol dependence[b]	1.2	1.0
Drug abuse[b]	1.4	1.4
Drug dependence[b]	1.0	1.0
Total		
Any disorder[b]	1.9	1.5

[a]Using R_weight_Part I.
[b]R_weight_Part II.

5 | CONCLUSION

This article presented an overview of the SNMHS sample design and weighting procedures. The SNMHS used a complex sample design that required weighting to adjust for differential probabilities of selection and the use of design-based estimation methods to correct for the under-estimation of standard errors with conventional methods due to the existence of clustering and the use of weighting. Importantly, different weights were created to account for different units of analysis that infer to different populations including Saudi households, Saudi citizens between the ages of 15 and 65, and Saudi citizens with disability. Different weights will consequently be used depending on the focus of analysis.

ACKNOWLEDGMENTS

The Saudi National Mental Health Survey (SNMHS) is conducted by the King Salman Center for Disability Research. It is funded by Saudi Basic Industries Corporation (SABIC), King Abdulaziz City for Science and Technology (KACST), Abraaj Capital, Ministry of Health (Saudi Arabia), and King Saud University. Funding in-kind was provided by King Faisal Specialist Hospital and Research Center, and the Ministry of Economy and Planning, General Authority for Statistics. None of the funders had any role in the design of the study, data analysis, interpretation of results, or preparation of this article. The SNMHS is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative. We thank the staff of the WMH Data Collection Coordination Centre in the Survey Research Center at University of Michigan and the WMH Data Analysis Coordination Centre in the Department of Health Care Policy at Harvard Medical School for assistance with design, instrumentation, fieldwork, and consultation on data analysis. A complete list of all WMH publications can be found at http://www.hcp.med.harvard.edu/wmh. We also acknowledge with gratitude the work and dedication of the SNMHS staff both current and past for their contributions to the study.

DECLARATION OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Zeina N. Mneimneh https://orcid.org/0000-0002-1091-7838

Yasmin A. Altwaijri https://orcid.org/0000-0001-8826-3224

REFERENCES

Al-Subaie, A. S., Al-Habeeb, A., Bilal, L., Shahab, M., Pennell, B.-E., Mneimneh, Z., … Altwaijri, Y. (In press). The Saudi National Mental Health Survey: Survey instrument and field procedures. *International Journal of Methods in Psychiatric Research*.

Alonso, J., Chatterji, S., & He, Y. (Eds.). (2013). *The burdens of mental disorders: Global perspectives from the WHO World Mental Health Surveys*. Cambridge, UK: Cambridge University Press.

American Association for Public Opinion Research. (2016). Standard definitions. Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

GBD 2015 Eastern Mediterranean Region Mental Health Collaborators. (2018). The burden of mental disorders in the eastern Mediterranean region, 1990-2015: Findings from the global burden of disease 2015 study. *International Journal of Public Health*, *63*(Suppl 1), 25–37. https://doi.org/10.1007/s00038-017-1006-1

General Authority for Statistics. (2010). *Population and housing characteristics in the Kingdom of Saudi Arabia: Demographic survey*. Riyadh, Saudi Arabia: General Authority for Statistics, Ministry of Economic and Planning.

Harkness, J., Pennell, B. E., Villar, A., Gebler, N., Aguilar-Gaxiola, S., & Bilgen, I. (2008). Translation procedures and translation assessment in the World Mental Health Survey Initiative. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health Surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge, UK: Cambridge University Press.

Heeringa, S. G., Wells, J. E., Hubbard, F., Mneimneh, Z. N., Chiu, W. T., Sampson, N. A., & Berglund, P. A. (2008). Sample designs and sampling procedures. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health Surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge, UK: Cambridge University Press.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). *Applied survey data analysis* (2nd ed.). Boca Raton, FL: Chapman and Hall.

Institute for Health Metrics and Evaluation. (2019). Saudi Arabia. Retrieved from http://www.healthdata.org/saudi-arabia

Kessler, R. C., Chatterji, S., Heeringa, S. G., Pennell, B.-E., Petukhova, M. V., Vilagut, G., & Zaslavsky, A. M. (2013). Methods of the World Mental Health Surveys. In J. Alonso, S. Chatterji, & Y. He (Eds.), *The burdens of mental disorders: Global perspectives from the World Mental Health Surveys*. Cambridge, UK: Cambrige University Press.

Kessler, R. C., Heeringa, S. G., Pennell, B.-E., Sampson, N. A., & Zaslavsky, A. M. (2018). Methods of the World Mental Health Surveys. In K. M. Scott, P. de Jonge, D. J. Stein, & R. C. Kessler (Eds.), *Mental disorders around the world: Facts and figures from the WHO World Mental Health Surveys*. Cambridge, UK: Cambrige University Press.

Kessler, R. C., & Üstün, T. B. (2008). *The WHO World Mental Health Surveys: Global perspective on the epidemiology of mental illness*. Cambridge, UK: Cambridge University Press.

Kish, L. (1965). *Survey sampling*. New York, NY: John Wiley & Sons.

Pennell, B. E., Mneimneh, Z. N., Bowers, A., Chardoul, S., Wells, J. E., Viana, M. C., … Saiz, G. V. (2008). Implementation of the World Mental Health Surveys. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO World Mental Health Surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge, UK: Cambridge University Press.

Scott, K. M., de Jonge, P., Stein, D. J., & Kessler, R. C. (Eds.). (2018). *Mental disorders around the world: Facts and figures from the WHO World Mental Health surveys*. Cambridge, UK: Cambridge University Press.

Wolter, K. M. (1985). *Introduction to variance estimation*. New York, NY: Springer-Verlag.

How to cite this article: Mneimneh ZN, Heeringa SG, Lin Y-C, Altwaijri YA, Nishimura R. The Saudi National Mental Health Survey: Sample design and weight development. *Int J Methods Psychiatr Res*. 2020;1–7. https://doi.org/10.1002/mpr.1829